MycoKeys 16: 1–15 (2016)
doi: 10.3897/mycokeys.16.10000
http://mycokeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

Annotating public fungal ITS sequences from the built environment according to the MIxS-Built Environment standard – a report from a May 23-24, 2016 workshop (Gothenburg, Sweden)

Kessy Abarenkov[1], Rachel I. Adams[2], Laszlo Irinyi[3,4,5], Ahto Agan[6,*], Elia Ambrosio[1,6,7,*], Alexandre Antonelli[8,9,*], Mohammad Bahram[6,10,*], Johan Bengtsson-Palme[11,*], Gunilla Bok[12,*], Patrik Cangren[8,*], Victor Coimbra[13,*], Claudia Coleine[14,*], Claes Gustafsson[15,*], Jinhong He[16,*], Tobias Hofmann[8,*], Erik Kristiansson[17,*], Ellen Larsson[8,*], Tomas Larsson[18,*], Yingkui Liu[8,*], Svante Martinsson[8,*], Wieland Meyer[3,4,5,*], Marina Panova[19,*], Nuttapon Pombubpa[20,*], Camila Ritter[8,*], Martin Ryberg[10,*], Sten Svantesson[8,*], Ruud Scharn[21,*], Ola Svensson[8,*], Mats Töpel[18,*], Martin Unterseher[22,*], Cobus Visagie[23,24,*], Christian Wurzbacher[8,*], Andy F. S. Taylor[25,28], Urmas Kõljalg[1,6], Lynn Schriml[26,27], R. Henrik Nilsson[8]

1 *University of Tartu, Natural History Museum, Tartu, Estonia* **2** *University of California, Plant & Microbial Biology, Berkeley, CA 94720, USA* **3** *Sydney Medical School-Westmead Hospital, Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Sydney, Australia* **4** *University of Sydney, Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney, Australia* **5** *Westmead Institute for Medical Research, Westmead, Australia* **6** *Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia* **7** *Via Calamandrei 2, 53035 Monteriggioni, Siena, Italy* **8** *University of Gothenburg, Department of Biological and Environmental Sciences, Box 461, 405 30 Göteborg, Sweden* **9** *Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, 413 19 Göteborg, Sweden* **10** *Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, SE 75236 Uppsala, Sweden* **11** *University of Gothenburg, Department of Infectious Diseases, Sahlgrenska Academy, Guldhedsgatan 10, 413 46 Göteborg, Sweden* **12** *SP Technical Research Institute of Sweden, Box 857, 501 15 Borås, Sweden* **13** *Universidade Federal de Pernambuco (UFPE), Departamento de Micologia, Centro de Ciências Biológicas (CCB), Av. Prof. Nelson Chaves, s/n, 50670-901 Recife, Pernambuco, Brazil* **14** *University of Tuscia, Department of Ecological and Biological Sciences, 01100, Viterbo, Italy* **15** *University of Gothenburg, Herbarium GB, Box 461, 405 30 Göteborg, Sweden* **16** *South China Botanical Garden, The Chinese Academy of Sciences, 723 Xingke Road, Tianhe District, Guangzhou 510650, China* **17** *Chalmers University of Technology, Department of Mathematical Sciences, 412 96 Göteborg, Sweden* **18** *University of Gothenburg, Department of Marine Sciences, Box 460, 405 30 Göteborg, Sweden* **19** *University of Gothenburg, Department of Marine Sciences - Tjärnö, 452 96, Strömstad, Sweden* **20** *University of California Riverside, Department of Plant Pathology and Microbiology and Institute of Integrative Genome Biology, Riverside, California 92521 USA* **21** *University of Gothenburg, Department of Earth*

* Co-authors contributed equally.

Copyright *Kessy Abarenkov et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Sciences, Box 460, 40530 Göteborg, Sweden **22** *Ernst-Moritz-Arndt University Greifswald, Institute of Botany and Landscape Ecology, Soldmannstr. 15, 17487 Greifswald, Germany* **23** *Agriculture and Agri-Food Canada, Biodiversity (Mycology), 960 Carling Avenue, Ottawa, ON Canada K1A 0C6* **24** *University of Ottawa, Department of Biology, 30 Marie-Curie, Ottawa, ON Canada K1N 6N5* **25** *The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland UK* **26** *Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA* **27** *Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA* **28** *Institute of Biological and Environmental Sciences, Cruickshank Building, University of Aberdeen, Aberdeen AB24 3UU, UK*

Corresponding author: *Kessy Abarenkov* (kessy.abarenkov@ut.ee)

Academic editor: *A. Miller* | Received 26 July 2016 | Accepted 8 September 2016 | Published 26 September 2016

Citation: Abarenkov K, Adams RI, Irinyi L, Agan A, Ambrosio E, Antonelli A, Bahram M, Bengtsson-Palme J, Bok G, Cangren P, Coimbra V, Coleine C, Gustafsson C, He J, Hofmann T, Kristiansson E, Larsson E, Larsson T, Liu Y, Martinsson S, Meyer W, Panova M, Pombubpa N, Ritter C, Ryberg M, Svantesson S, Scharn R, Svensson O, Töpel M, Unterseher M, Visagie C, Wurzbacher C, Taylor AFS, Kõljalg U, Schriml L, Nilsson RH (2016) Annotating public fungal ITS sequences from the built environment according to the MIxS-Built Environment standard – a report from a May 23-24, 2016 workshop (Gothenburg, Sweden). MycoKeys 16: 1–15. doi: 10.3897/mycokeys.16.10000

Abstract

Recent molecular studies have identified substantial fungal diversity in indoor environments. Fungi and fungal particles have been linked to a range of potentially unwanted effects in the built environment, including asthma, decay of building materials, and food spoilage. The study of the built mycobiome is hampered by a number of constraints, one of which is the poor state of the metadata annotation of fungal DNA sequences from the built environment in public databases. In order to enable precise interrogation of such data – for example, "retrieve all fungal sequences recovered from bathrooms" – a workshop was organized at the University of Gothenburg (May 23–24, 2016) to annotate public fungal barcode (ITS) sequences according to the MIxS-Built Environment annotation standard (http://gensc.org/mixs/). The 36 participants assembled a total of 45,488 data points from the published literature, including the addition of 8,430 instances of countries of collection from a total of 83 countries, 5,801 instances of building types, and 3,876 instances of surface-air contaminants. The results were implemented in the UNITE database for molecular identification of fungi (http://unite.ut.ee) and were shared with other online resources. Data obtained from human/animal pathogenic fungi will furthermore be verified on culture based metadata for subsequent inclusion in the ISHAM-ITS database (http://its.mycologylab.org).

Key words

Built environment, Indoor fungi, ITS, Annotation, Mycobiome

Introduction

Fungi are found throughout the biosphere, and the built environment is no exception. The taxonomic composition of indoor fungal communities tends to reflect the local outdoor communities, although the majority of fungal particles found indoors is

thought to represent spores, hyphal fragments, and other dormant and passively distributed stages (Seo et al. 2015). Although most of the fungi recovered from indoor environments would not be able to live in the built environment for any extended period of time, a minority of these species are able to cope with, and will even thrive in, the harsh conditions that the built environment presents (Hamada and Abe 2010; Nevalainen et al. 2015; Zupančič et al. 2016). These species are mainly saprotrophic, and their degree of active growth largely depends on water availability (Adams et al. 2013). They can be a serious cause of decay and other concerns in water-damaged buildings, but they are also found in buildings not subject to moisture issues – and even in buildings where very strict sanitization and filtration regimes are applied (e.g., La Duc et al. 2012; Checinska et al. 2015). Exposure to aerosolized fungal particles has been linked to asthma onset in humans and may furthermore play a role in eczema development and other issues in human health (Reijula et al. 2003; Knutsen et al. 2012). Indoor fungi may also contribute to other unwanted processes, such as food spoilage and wall staining (Varga et al. 2014). The built mycobiome is thus of interest to a range of scientific fields, including mycology, medicine, food biology, construction, and engineering.

Traditional, morphology-based studies of fungal spores and cultures derived from indoor sampling have recognized ca. 90 species of common indoor fungi (Flannigan et al. 2002). Efforts based on high-throughput DNA sequencing, in contrast, have revealed a vast and hitherto unknown diversity of indoor fungi. In a global study of indoor dust samples, Amend et al. (2010) using next-generation sequencing found ca. 4,500 fungal operational taxonomic units (OTUs; Blaxter et al. 2005) approximately at the species level. Similarly, another next-generation sequencing-powered study – Nonnenmann et al. (2012) – recovered 450 fungal species from 50 indoor dust samples in Yakima valley, WA (USA). Although precise species delimitation and species counts from next-generation sequencing data remain challenging (Nguyen et al. 2016), the taxonomic span of the fungal assemblages recovered in Amend et al. (2010) and Nonnenmann et al. (2012) is far larger than that occupied by the fungi traditionally thought of as common indoor fungi (cf. Flannigan et al. 2002). Thus, whereas these studies should not be used as estimates of the total number of indoor fungi, they do testify to the substantial diversity of fungi in the built environment. The lack of taxonomic reference sequences makes precise identification of many of these species problematic, and it is not unusual that a sizable proportion of the OTUs in environmental sequencing studies remain unassigned beyond the kingdom or phylum levels (e.g., Tedersoo et al. 2014; Grossart et al. 2016; Fouquier et al. 2016; Nilsson et al. 2016). There is clearly a need to generate reliable reference sequences, most notably from type material, to address this issue (cf. Schoch et al. 2014). However, the estimated number of extant species of fungi – 1.5-6 million (Hawksworth 2001; Taylor et al. 2014) – stands in stark contrast to the number of described species (~130,000 as of March 2016; www.speciesfungorum.org), and strongly suggests that molecular identification of fungi will remain challenging for the foreseeable future. In some cases, even reference barcode (nuclear

ribosomal internal transcribed spacer, ITS) sequences from type material will not be enough. Several fungal genera regularly recovered from built environment samples - such as *Aspergillus*, *Cladosporium*, *Fusarium*, and *Penicillium* - show little or no ITS variation across sets of two to several species (Bensch et al. 2012; Samson et al. 2014; Visagie et al. 2014; O'Donnell et al. 2015). Additional genetic markers are needed for robust species-level identification in these cases.

A second problem that compounds the scientific understanding of the built mycobiome has been the lack of a standardized vocabulary for sequence annotation. The International Nucleotide Sequence Database Collaboration (INSDC; Cochrane et al. 2016) holds more than 5,000 Sanger-derived fungal ITS (barcode) sequences from the built environment, but their level of metadata annotation differs widely. This unfortunately applies to most available fungal ITS sequences (cf. Nilsson et al. 2014); for example, a modest 43% are known to be annotated with something as simple and straightforward as country of collection (Tedersoo et al. 2011). In addition, where metadata exist they are not always provided in standardized and searchable formats, making precise queries difficult. There is, for instance, no straightforward way to download all fungal ITS sequences from bathrooms, or to target the substrate of gypsum board. It is reasonable to think that analysis of fungi recovered from bathrooms may prove a rewarding scientific enterprise, as indeed should be the case for fungi collected on specific building materials, under different moisture regimes, or any other particular parameter or setting. The full potential of such searches cannot presently be utilized due to the poor state of sequence annotation – primarily omitted by the original sequence authors – in the public sequence databases.

The new MIxS-Built Environment annotation standard (Glass et al. 2014; http://gensc.org/mixs/) addresses the need for a thorough, standardized vocabulary for microbiological analysis of the built microbiome. If all relevant fungal ITS sequences in the INSDC were annotated according to this standard, then this would open up the body of extant molecular data to detailed, precise scientific queries in the context of the built mycobiome. Going through and annotating large sequence sets is a daunting effort for any researcher, but fortunately such efforts are easy to split among a set of individual researchers. This paper presents the outcome of a sequence metadata annotation workshop (University of Gothenburg, May 23-24, 2016) to annotate the ~6,500 public fungal ITS sequences from the built environment according to the most relevant parts of the MIxS-Built Environment annotation standard. In recognition of the fact that fungi found indoors are typically found outdoors as well, the workshop also annotated closely related outdoor sequences according to basic geo-ecological parameters. The workshop was organized jointly with the UNITE and ISHAM databases (Kõljalg et al. 2013; Irinyi et al. 2015). UNITE is a general-purpose sequence management environment seeking to reconcile molecular ecology and taxonomy of fungi and fungal communities. The ISHAM database centers on identification of human and animal pathogenic fungi to guide antifungal treatment choices. Both databases focus, at least for the time being, on the ITS region and share views on the importance of openness, free accessibility, and community participation.

Materials and methods

The workshop comprised 20 physical participants, mainly local Ph.D. students and postdocs – but also other researchers – in systematics and ecology. In addition, another 16 researchers participated remotely through Skype, Google Docs, and email. The participants focused on the public fungal ITS sequences of the INSDC as mirrored in the UNITE and ISHAM databases. To single out INSDC sequences associated with the built environment, we used a set of 24 keywords such as "dust", "gypsum", and "floor" (Suppl. material 1). Keyword matches were made to the title of the underlying publication (the INSDC field "title"), the INSDC fields "source" and "tissue type", and the UNITE field "sequence source". We refer to this set of sequences as the *built mycobiome set* (BMS). To single out outdoor sequences with a direct relation to the BMS, we extracted all UNITE species hypotheses with at least one BMS sequence. We then built the *outdoor mycobiome set* (OMS) from all sequences that did not match any of our keywords but that were found in the same species hypothesis as at least one BMS sequence. Sequences that initially were assigned to the BMS set, but that on closer inspection turned out not to qualify as the built mycobiome ("collected outside hospital", for example), were transferred to the OMS set.

For each BMS sequence we tried to locate any underlying publication through the INSDC fields TITLE, JOURNAL, and PUBMED. If these were not informative, we resorted to ISI Thompson, Google/Google Scholar, and ResearchGate searches. We examined the publications for the nine items of the MIxS-Built Environment annotation standard that we felt were the most relevant and the most likely to be covered by the studies: building occupancy type, indoor space, indoor surface, surface material, surface-air contaminant, space typical state, substructure type, ventilation type, and filter type (http://gensc.org/mixs/). In addition we also targeted the country and host of collection and the nature of the fungus-host association (e.g., "plant: wood", "plant: leaf", and "human/animal: skin"), as applicable, for all sequences. We only targeted metadata and information that was clearly and unequivocally specified in the paper. A research professional (G. Bok) from a building-related technical institute was present to assist with technical, analytical, and construction-related questions in the context of the built environment. For the OMS we similarly retrieved the underlying publications and annotated the sequences to country and host of collection plus host association (as applicable, and if and when these data were missing). All results were entered into an Excel sheet for upload into UNITE and ISHAM (after culture-based verification in the case of the latter), and for sharing with other online resources.

Results

A total of 6,526 BMS and 11,574 OMS sequences from a total of 255 separate studies were annotated with at least one metadata item. A total of 45,488 annotations were made during the workshop. For example, "building occupancy type" was established for 5,801 sequences, and "ventilation type" was established for 2,235 sequences (Table 1;

Table 1. Results of the annotation workshop, specified for the built mycobiome sequence set (BMS) and the outdoor mycobiome sequence set (OMS). Countries and hosts of collections plus host association were assembled for both of these. The number of sequences processed, plus the number of underlying published and unpublished scientific studies, are also provided. For the BMS, the nine MIxS-Built Environment annotation standard items targeted at the workshop are specified in separate columns. The sequence numbers shown in the table refer to the number of sequences annotated for each data item.

	Number of sequences (annotated)	**Number of studies**	**Country of collection**	**Different countries**	**Host of collection**	**Different hosts**
BMS	6550 (6526)	144	2447	29	881	15
OMS	16766 (11574)	128	5983	83	5632	859
Total	23316 (18100)	255 unique	8430	83 unique	6513	865 unique
	Host association	**Comments**	**Building occupancy type**	**Indoor space**	**Indoor surface**	**Surface material**
BMS	764	2348	5801	1223	1207	1318
OMS	2892	1293	N/A	N/A	N/A	N/A
Total	3656	3641	5801	1223	1207	1318
	Surface-air contaminant	**Space typical state**	**Substructure type**	**Ventilation type**	**Filter type**	
BMS	3876	5618	96	2235	1874	
OMS	N/A	N/A	N/A	N/A	N/A	
Total	3876	5618	96	2235	1874	

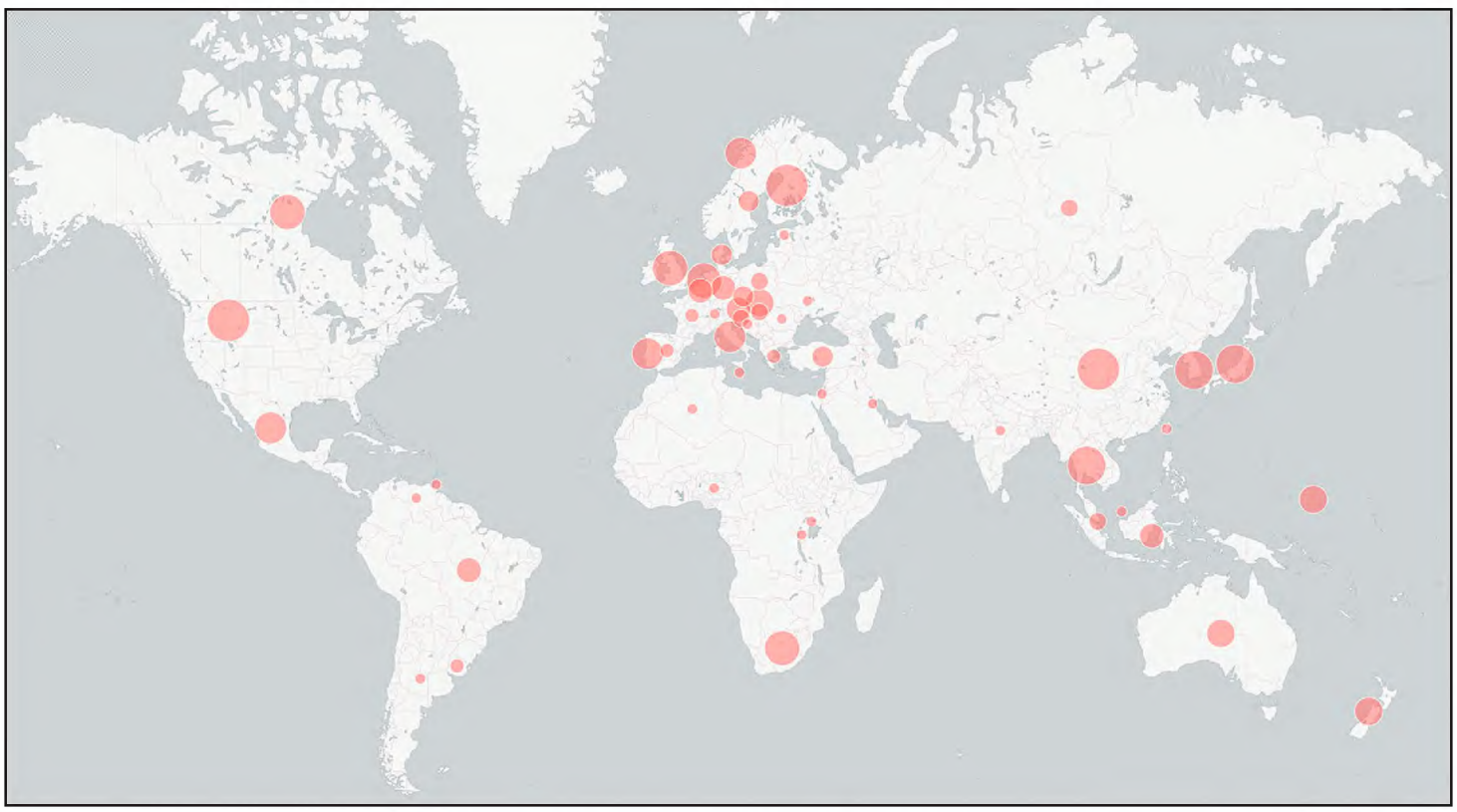

Figure 1. Analysis of the BMS sequences for country of collection. Country centroids marked with bubbles of different size on the global map indicate the number of BMS sequences originating from these countries (54 distinct countries, sequence count ranging from 1 to 2,914). For an additional 2.9% of the sequences, country information could not be restored during the workshop. The figure includes pre-existing data plus the data added during the workshop, such that these charts indicate the scientific state of ITS-based Sanger-derived sequencing of the built mycobiome as of spring 2016. Sequences that were not annotated with a single built environment-related term in the INSDC were not included in this effort, and are not represented in these charts.

Figures 1–3). The results were uploaded into UNITE via its data management system PlutoF (https://plutof.ut.ee; Abarenkov et al. 2010) for open query by the scientific community and was shared with the INSDC as an Excel sheet (Suppl. material 2).

Discussion

The workshop compiled a total of 45,488 metadata items, making them available for scientific query through UNITE and other venues. These metadata, although typically "published" and thus "available", were previously not open for direct query. This highlights the wealth of relevant scientific information that lies buried in the last few decades' worth of scientific publications – formally available, yet only available to those who know where to look, and reachable only to those with access to that literature. Fortunately, we live in a digital age where the infrastructure for recovering and sharing such information is falling into place (Martin and Martin 2010). Furthermore, there is a growing awareness of the need to annotate newly generated sequences beyond the barest minimum when these are first deposited into public sequence databases (Hyde et al. 2013; Schoch et al. 2014). Such annotations unlock significant scientific potential of those molecular data, increase the citability of the underlying scientific studies, and

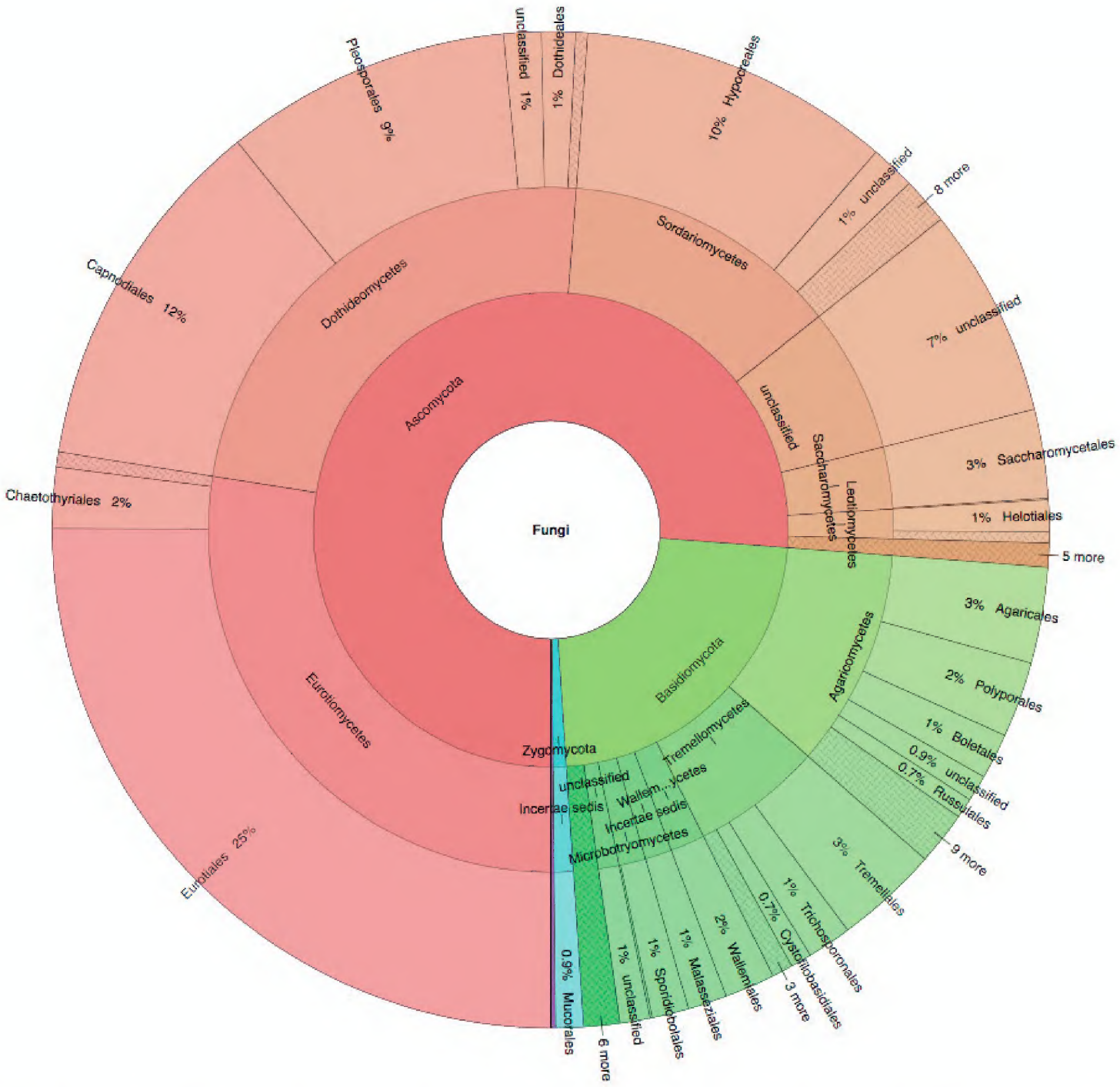

Figure 2. Krona chart of the taxonomic affiliation of the BMS sequences down to order level. The Krona chart lists all annotated BMS sequences except those classified as Fungi sp. (36.4%) and those of non-fungal origin (0.9%). An interactive version of the Krona chart is provided as Suppl. material 3. The figure includes pre-existing data plus the data added during the workshop, such that these charts indicate the scientific state of ITS-based Sanger-derived sequencing of the built mycobiome as of spring 2016. Sequences that were not annotated with a single built environment-related term in the INSDC were not included in this effort, and are not represented in these charts.

fulfill funding agencies' demands for openness and maximum scientific use of research funding. We certainly hope that the mycological community will be quick to embrace a more integrative approach to sequence annotation. The public sequence databases can similarly make it even easier and faster to provide such metadata upon sequence submission. We speculate that excessive time consumption is the primary reason why some sequence depositors do not annotate their sequences as well as they could have.

We managed to process nearly all BMS sequences – for which we could retrieve the underlying publication(s) – for at least one metadata item. A total of 4,985 sequences

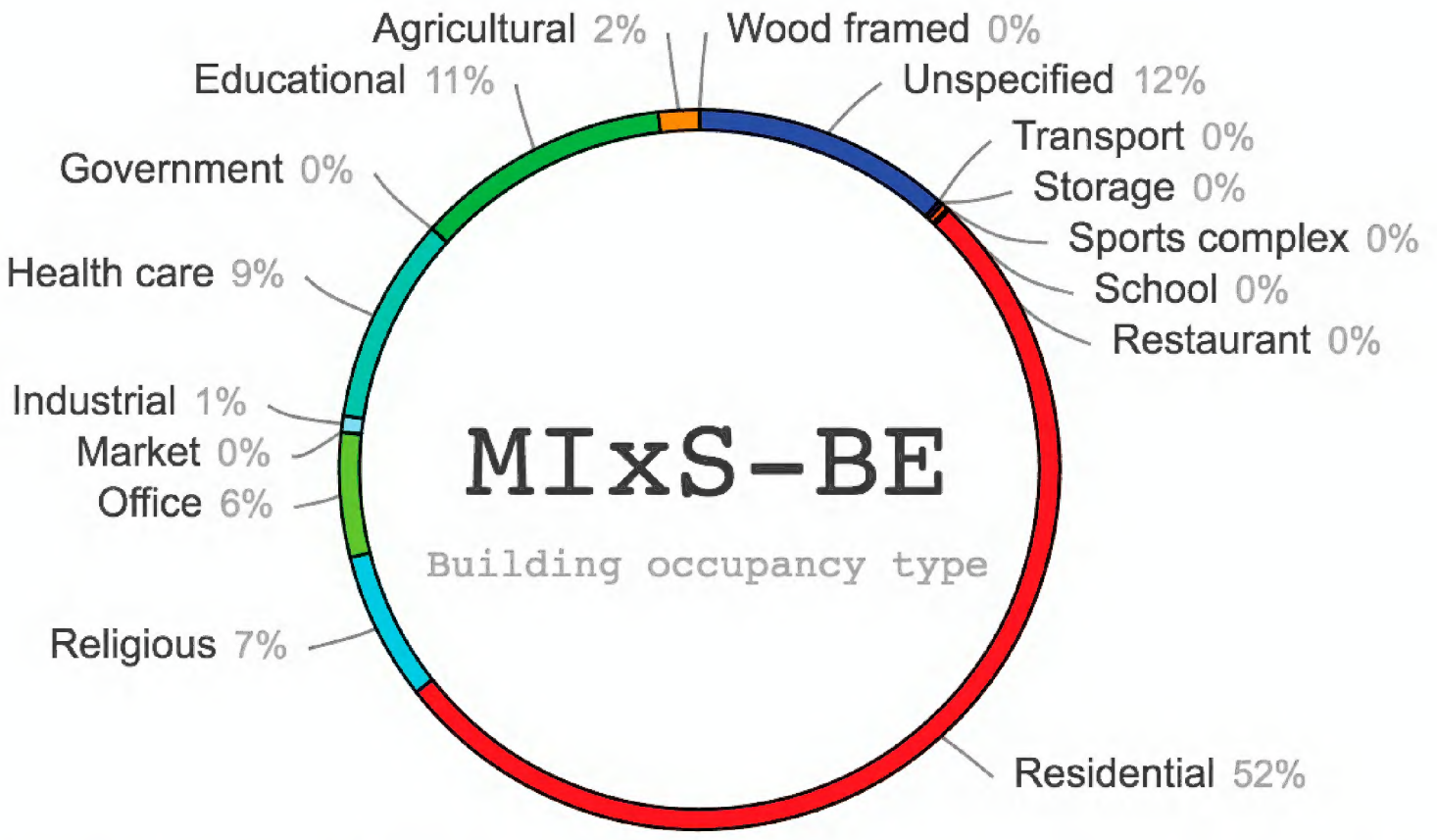

Figure 3. Analysis of the MIxS-BE "building occupancy type" (type of building where the underlying sample was taken).

were false positives – our keywords indicated them to belong to the BMS whereas in reality they did not. A sequence could stem from "outside city hospital" (keyword "hospital"), for instance. These sequences were annotated for country and host of sampling, plus the nature of the relation to the host, whenever the underlying scientific study could be retrieved and interpreted. It is reasonable to assume that our initiative suffered from a fair number of false negatives as well – sequences that should have been a part of the BMS, but that were not. Although we used no fewer than 24 keywords in our efforts to capture the built environment, we presumably missed one or more important terms in the field. We similarly missed out on all built-environment sequences that featured no relevant annotation whatsoever – perhaps just a species name and the country of origin were available. Thus, whereas we managed to do at least something about nearly all BMS sequences we recovered, we do not claim to have annotated all public fungal ITS sequences from the built environment.

The workshop identified several potential venues for amendments to the MIxS-BE standard. For example, "floor" was found to be a common place for sampling of, e.g., dust, yet the data point of "floor" could not easily be fitted into any extant MIxS-BE category. Similarly, "air" could not be represented in a straightforward way in the MIxS-BE standard (but rather applied to other packages of the MIxS standard). We also felt the need for a "laboratory" flag to indicate that a sequence stemmed from sampling in a laboratory. In addition, we were surprised by the number of fungal sequences generated from environments that must be considered to qualify as "built" or at least altered by man, but that nevertheless were difficult to fit into the present MIxS-BE categories. The examples included tombs, crypts, and mummies (Šimonovičová et

al. 2015), tumuli and other prehistoric remains (e.g., Kiyuna et al. 2011; Fernandez-Cortes et al. 2011), spacecraft (Satoh et al. 2011), and indoor historical paintings or artifacts such as the Turin shroud (López-Miras et al. 2013; Barcaccia et al. 2015). In these cases, we tried to capture the essence of the underlying sequence entries to the extent that the MIxS-BE standard allowed. We used our free-text field "Comment" to provide additional information that we felt was important with respect to future queries of these entries. These potential venues for improvements of the MIxS-BE standard have been communicated to MIxS-BE representatives from the Genomic Standards Consortium's MIxS Compliance and Implementation working group (http://gensc.org/mixs/mixs-compliance-and-implementation/).

Conclusions

The present study used a workshop-style approach to accomplish a task that would have taken several months for a single researcher to accomplish. Costs were kept low by recruiting many of the participants among local Ph.D. students and postdocs in systematics and ecology, and workshop participation was made attractive by providing the opportunity to contribute to this workshop report. We can recommend this model when tackling projects of a similar kind, such as data assembly and analysis in molecular ecology and systematics. As an added benefit, the more junior participants obtain experience in scientific collaboration and communication as well as in carrying out scientific projects (cf. Ryberg et al. 2016). The workshop was funded by an Alfred P. Sloan foundation grant to improve the support for the built mycobiome in UNITE and elsewhere. Other events include a forthcoming (2017) taxonomic sequence annotation workshop and the generation and public release of sequences from type material. We invite feedback and participation in these events, and we welcome any other idea to take molecular identification of the built mycobiome to the next level.

Acknowledgements

We gratefully acknowledge financial support from the Alfred P. Sloan Foundation and the Swedish Research Council of Environment, Agricultural Sciences, and Spatial Planning (FORMAS, 215-2011-498). VRMC thanks CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (SWE 232695/2014-8) for providing a shared Ph.D. scholarship. A.A. is funded by the Swedish Research Council (B0569601), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement n. 331024), and a Wallenberg Academy Fellowship.

References

Abarenkov K, Tedersoo L, Nilsson RH et al. (2010) PlutoF - a web-based workbench for ecological and taxonomical research, with an online implementation for fungal ITS sequences. Evolutionary Bioinformatics 6: 189–196. doi: 10.4137/EBO.S6271

Adams RI, Miletto M, Taylor JW, Bruns TD (2013) Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. The ISME Journal 7(7): 1262–1273. doi: 10.1038/ismej.2013.28

Amend AS, Seifert KA, Samson R, Bruns TD (2010) Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. Proceedings of the National Academy of Sciences 107(31): 13748–3753. doi: 10.1073/pnas.1000454107

Barcaccia G, Galla G, Achilli A, Olivieri A, Torroni A (2015) DNA analysis of dust particles sampled from the Turin Shroud. MATEC Web of Conferences 36: 03001. doi: 10.1051/matecconf/20153603001

Bensch K, Braun U, Groenewald JZ, Crous PW (2012) The genus *Cladosporium*. Studies in Mycology 72: 1–401. doi: 10.3114/sim0003

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360(1462): 1935–1943. doi: 10.1098/rstb.2005.1725

Checinska A, Probst AJ, Vaishampayan P, White JR, Kumar D, Stepanov VG, Fox GE, Nilsson RH, Pierson DL, Perry J, Venkateswaran K (2015) Microbiomes of the dust particles collected from the International Space Station and Spacecraft Assembly Facilities. Microbiome 3(1): 1. doi: 10.1186/s40168-015-0116-3

Cochrane G, Karsch-Mizrachi I, Takagi T, International Nucleotide Sequence Database Collaboration (2016) The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research 44(D1): D48–D50. doi: 10.1093/nar/gkv1323

Fernandez-Cortes A, Cuezva S, Sanchez-Moral S, Cañaveras JC, Porca E, Jurado V, Martin-Sanchez PM, Saiz-Jimenez C (2011) Detection of human-induced environmental disturbances in a show cave. Environmental Science and Pollution Research 18(6): 1037–1045. doi: 10.1007/s11356-011-0513-5

Flannigan B, Samson RA, Miller JD (2002) Microorganisms in home and indoor work environments, 2nd edition. CRC Press.

Fouquier J, Schwartz T, Kelley ST (2016) Rapid assemblage of diverse environmental fungal communities on public restroom floors. Indoor air: In press. doi: 10.1111/ina.12279

Glass EM, Dribinsky Y, Yilmaz P et al. (2014) MIxS-BE: a MIxS extension defining a minimum information standard for sequence data from the built environment. ISME Journal 8(1): 1–3. doi: 10.1038/ismej.2013.176

Grossart HP, Wurzbacher C, James TY, Kagami M (2016) Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. Fungal Ecology 19: 28–38. doi: 10.1016/j.funeco.2015.06.004

Hamada N, Abe N (2010) Comparison of fungi found in bathrooms and sinks. Biocontrol Science 15(2): 51–56. doi: 10.4265/bio.15.51

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycological Research 105(12): 1422–1432. doi: 10.1017/S0953756201004725

Hyde KD, Udayanga D, Manamgoda DS et al. (2013) Incorporating molecular data in fungal systematics: a guide for aspiring researchers. Current Research in Environmental and Applied Mycology 3(1): 1–32. doi: 10.5943/cream/3/1/1

Irinyi L, Serena C, Garcia-Hermoso D et al. (2015) International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database -the quality controlled standard tool for routine identification of human and animal pathogenic fungi. Medical Mycology 53(4): 313–337. doi: 10.1093/mmy/myv008

Kiyuna T, An KD, Kigawa R, Sano C, Miura S, Sugiyama J (2011) Molecular assessment of fungi in "black spots" that deface murals in the Takamatsuzuka and Kitora Tumuli in Japan: *Acremonium* sect. *Gliomastix* including *Acremonium tumulicola* sp. nov. and *Acremonium felinum* comb. nov. Mycoscience 52(1): 1–17. doi: 10.1007/s10267-010-0063-6

Knutsen AP, Bush RK, Demain JG et al. (2012) Fungi and allergic lower respiratory tract diseases. Journal of Allergy and Clinical Immunology 129(2): 280–291. doi: 10.1016/j.jaci.2011.12.970

Kõljalg U, Nilsson RH, Abarenkov K et al. (2013) Towards a unified paradigm for sequence-based identification of Fungi. Molecular Ecology 22(21): 5271–5277. doi: 10.1111/mec.12481

La Duc MT, Vaishampayan P, Nilsson RH, Torok T, Venkateswaran K (2012) Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. Applied and Environmental Microbiology 78(16): 5912–5922. doi: 10.1128/AEM.01435-12

López-Miras M, Piñar G, Romero-Noguera J, Bolivar-Galiano FC, Ettenauer J, Sterflinger K, Martín-Sánchez I (2013) Microbial communities adhering to the obverse and reverse sides of an oil painting on canvas: identification and evaluation of their biodegradative potential. Aerobiologia 29(2): 301–314. doi: 10.1007/s10453-012-9281-z

Martin NF, Martin F (2010) From galactic archeology to soil metagenomics –surfing on massive data streams. New Phytologist 185(2): 343–347. doi: 10.1111/j.1469-8137.2009.03138.x

Nevalainen A, Täubel M, Hyvärinen A (2015) Indoor fungi: companions and contaminants. Indoor Air 25(2): 125–156. doi:10.1111/ina.12182

Nguyen NP, Warnow T, Pop M, White B (2016) A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. npj Biofilms and Microbiomes 2: 16004. doi: 10.1038/npjbiofilms.2016.4

Nilsson RH, Hyde KD, Pawlowska J et al. (2014) Improving ITS sequence data for identification of plant pathogenic fungi. Fungal Diversity 67(1): 11–19. doi: 10.1007/s13225-014-0291-8

Nilsson RH, Wurzbacher C, Bahram M et al. (2016) Top 50 most wanted fungi. MycoKeys 12: 29–40. doi: 10.3897/mycokeys.12.7553

Nonnenmann MW, Coronado G, Thompson B, Griffith WC, Hanson JD, Vesper S, Faustman EM (2012) Utilizing pyrosequencing and quantitative PCR to characterize fungal populations among house dust samples. Journal of Environmental Monitoring 14(8): 2038–2043. doi: 10.1039/c2em30229b

O'Donnell K, Ward TJ, Robert VARG, Crous PW, Geiser DM, Kang S (2015) DNA sequence-based identification of *Fusarium*: current status and future directions. Phytoparasitica 43(5): 583–595. doi: 10.1007/s12600-015-0484-z

Reijula K, Leino M, Mussalo-Rauhamaa H, Nikulin M et al. (2003) IgE-mediated allergy to fungal allergens in Finland with special reference to *Alternaria alternata* and *Cladosporium herbarum*. Annals of Allergy, Asthma & Immunology 91(3): 280–287. doi: 10.1016/S1081-1206(10)63531-4

Ryberg M, Kristiansson E, Wurzbacher C, Nilsson RH (2016) New Ph.D. students in the empirical sciences should be recruited into ongoing scientific studies right from the start. Journal of Brief Ideas. doi 10.5281/zenodo.48158

Samson RA, Visagie CM, Houbraken J et al. (2014) Phylogeny, identification and nomenclature of the genus *Aspergillus*. Studies in Mycology 78: 141–173. doi: 10.1016/j.simyco.2014.07.004

Satoh K, Nishiyama Y, Yamazaki T et al. (2011) Microbe-I: fungal biota analyses of the Japanese experimental module KIBO of the International Space Station before launch and after being in orbit for about 460 days. Microbiology and Immunology 55(12): 823–829. doi: 10.1111/j.1348-0421.2011.00386.x

Schoch CL, Robbertse B, Robert V et al. (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. Database, 2014, bau061. doi: 10.1093/database/bau061

Seo S, Ji YG, Yoo Y, Kwon MH, Choung JT (2015) Submicron fungal fragments as another indoor biocontaminant in elementary schools. Environmental Science: Processes & Impacts 17(6): 1164–1172. doi: 10.1039/C4EM00702F

Šimonovičová A, Kraková L, Pangallo D, Majorošová M, Piecková E, Bodoriková S, Dörnhoferová M (2015) Fungi on mummified human remains and in the indoor air in the Kuffner family crypt in Sládkovičovo (Slovakia). International Biodeterioration & Biodegradation 99: 157–164. doi: 10.1016/j.ibiod.2014.12.011

Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, Ruess RW (2014) A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. Ecological Monographs 84(1): 3–20. doi: 10.1890/12-1693.1

Tedersoo L, Abarenkov K, Nilsson RH et al. (2011) Tidying up International Nucleotide Sequence Databases: ecological, geographical, and sequence quality annotation of ITS sequences of mycorrhizal fungi. PLoS ONE 6(9): e24940. doi: 10.1371/journal.pone.0024940

Tedersoo L, Bahram M, Põlme S et al. (2014) Global diversity and geography of soil fungi. Science 346(6213): 1256688. doi: 10.1126/science.1256688

Varga J, Kocsubé S, Szigeti G, Baranyi N et al. (2014) Occurrence of black *Aspergilli* in indoor environments of six countries. Archives of Industrial Hygiene and Toxicology 65(2): 219–223. doi: 10.2478/10004-1254-65-2014-2450

Visagie CM, Houbraken J, Frisvad JC, Hong SB, Klaassen CHW, Perrone G, Seifert KA, Varga J, Samson RA (2014) Identification and nomenclature of the genus *Penicillium*. Studies in Mycology 78: 343–371. doi: 10.1016/j.simyco.2014.09.001

Zupančič J, Babič MN, Zalar P, Gunde-Cimerman N (2016) The black yeast *Exophiala dermatitidis* and other selected opportunistic human fungal pathogens spread from dishwashers to kitchens. PLoS ONE 11(2): e0148166. doi: 10.1371/journal.pone.0148166

Supplementary material 1

Keywords used to identify fungal sequences from the built environment in the INSDC

Authors: Kessy Abarenkov Rachel I. Adams Irinyi Laszlo Ahto Agan, Elia Ambrosio, Alexandre Antonelli, Mohammad Bahram, Johan Bengtsson-Palme, Gunilla Bok, Patrik Cangren, Victor Coimbra, Claudia Coleine, Claes Gustafsson, Jinhong He, Tobias Hofmann, Erik Kristiansson, Ellen Larsson, Tomas Larsson, Yingkui Liu, Svante Martinsson, Wieland Meyer, Marina Panova, Nuttapon Pombubpa, Camila Ritter, Martin Ryberg, Sten Svantesson, Ruud Scharn, Ola Svensson, Mats Töpel, Martin Unterseher, Cobus Visagie, Christian Wurzbacher, Andy F.S. Taylor Urmas Kõljalg Lynn Schriml R. Henrik Nilsson

Data type: text

Explanation note: Keywords used to identify fungal sequences from the built environment in the INSDC.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 2

Annotations made during the workshop

Authors: Kessy Abarenkov Rachel I. Adams Irinyi Laszlo Ahto Agan, Elia Ambrosio, Alexandre Antonelli, Mohammad Bahram, Johan Bengtsson-Palme, Gunilla Bok, Patrik Cangren, Victor Coimbra, Claudia Coleine, Claes Gustafsson, Jinhong He, Tobias Hofmann, Erik Kristiansson, Ellen Larsson, Tomas Larsson, Yingkui Liu, Svante Martinsson, Wieland Meyer, Marina Panova, Nuttapon Pombubpa, Camila Ritter, Martin Ryberg, Sten Svantesson, Ruud Scharn, Ola Svensson, Mats Töpel, Martin Unterseher, Cobus Visagie, Christian Wurzbacher, Andy F.S. Taylor Urmas Kõljalg Lynn Schriml R. Henrik Nilsson

Data type: metadata

Explanation note: The annotations made during the workshop shown with original INSDC data. For the BMS, we targeted nine MIxS-BE items plus country of collection, host of collection, host association, and a general "Comment" field. For the OMS, we targeted country of collection, host of collection, host association, and a general "Comment" field.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 3

Krona chart

Authors: Kessy Abarenkov Rachel I. Adams Irinyi Laszlo Ahto Agan, Elia Ambrosio, Alexandre Antonelli, Mohammad Bahram, Johan Bengtsson-Palme, Gunilla Bok, Patrik Cangren, Victor Coimbra, Claudia Coleine, Claes Gustafsson, Jinhong He, Tobias Hofmann, Erik Kristiansson, Ellen Larsson, Tomas Larsson, Yingkui Liu, Svante Martinsson, Wieland Meyer, Marina Panova, Nuttapon Pombubpa, Camila Ritter, Martin Ryberg, Sten Svantesson, Ruud Scharn, Ola Svensson, Mats Töpel, Martin Unterseher, Cobus Visagie, Christian Wurzbacher, Andy F.S. Taylor Urmas Kõljalg Lynn Schriml R. Henrik Nilsson

Data type: html

Explanation note: Interactive Krona chart for visualizing the taxonomic distribution of annotated BMS sequences down to order level. Sequences classified as Fungi sp. (36.4%) or non-fungal (0.9%) were excluded from this dataset.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.